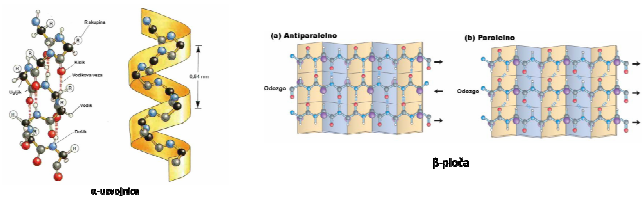


1 Uvod

Proteini imaju ključnu ulogu u gotovo svim biološkim procesima. Ta uloga definirana je funkcijom proteina koju određuje njegova struktura. Funkcija, struktura i interakcije proteina pripadnih sljedova uglavnom su nepoznate. Eksperimentalne metode veoma su skupe i vremenski zahtjevne pa je sve veći nesrazmjer između broja poznatih sljedova i pripadnih struktura. Stoga se sve veći naglasak stavlja na stvaranje i unaprjeđivanje računskih metoda koje će na temelju poznatih aminokiselinskih nizova vjerodostojno predvidjeti konačnu strukturu proteina.

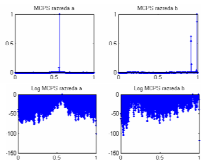


2 Opis problema

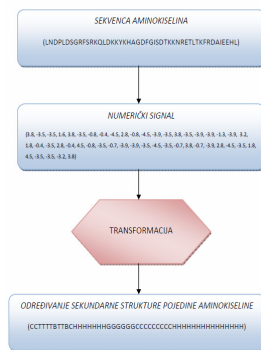
Cilj rada jest predvidjeti sekundarne strukture i strukturne razrede sekvenci aminokiselina proteinskih lanaca pomoću metoda obrade signala. Na temelju niza aminokiselina (primarna struktura) stvaraju se signali nad kojima se vrši obrada kako bi se izlučila značajka koja bi pomogla u određivanju sekundarne strukture proteina.

3 Metoda

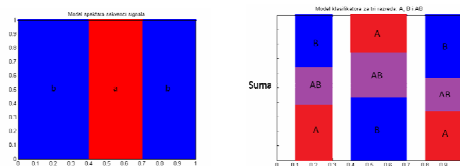
Svakoj aminokiselini sekvence pridijeljena je njena vrijednost indeksa hidrofobnosti čime je stvoren diskretni signal pogodan za obradu. Na temelju amplitudnog spektra te njegovih mjera, pokušavaju se uvidjeti karakteristične spektralne komponente za pojedini spektralni razred, odnosno tip sekundarne strukture.



Tipovi strukturnih razreda	
a	Metodski zastupljeni α-izvojnici
b	Metodski zastupljeni β-ploče
c	Izoprime pilače α-izvojnici β-ploče
d	Striktno odvojene α-izvojnici β-ploče



Na temelju amplitudnih spektara predloženi su modeli spektara signala sekvenci. Modeli predstavljaju očekivanu raspodjelu energije pojedinih dijelova spektra za pojedine strukturne razrede.



Modeli sekvenci za uključaj dva, odnosno tri spektralna razreda

Također, nad amplitudnim spektrima provedena je analiza glavnih komponenti te su sekvence klasificirane pomoću SVM (*support vector machine*) klasifikatora.

4 Rezultati

Najbolje poklapanje modela sekvenci strukturnih razreda *a* i *b* dobiveno je za parametre prikazane u sljedećoj tablici:

Frekv. područje razreda a	Frekv. područje razreda b	Vrijednost praga (threshold)
$(0.3\pi, 0.7\pi]$	$[0, 0.3\pi] \cup (0.7\pi, \pi]$	0.4
Točnost na skupu a	Točnost na skupu b	Ukupna točnost
66%	92%	83%

Pretpostavka o raspodjeli energije po frekvencijskim područjima za spektralne razrede *a* i *b* pokazala se relativno točnom. Korištenjem SVM klasifikatora dobiveni su sljedeći rezultati:

Klasifikacija razreda a i b		
Širina Gaussove funkcije	Točnost na a	Točnost na b
100	70%	82%

Klasifikacija razreda a i b (korišteno prvih pet glavnih komponenti)		
Širina Gaussove funkcije	Točnost na a	Točnost na b
10	74%	76%

Skoro pa jednaki rezultati dobivaju se koristeći samo prvih pet glavnih spektralnih komponenti što znači da one skoro u potpunosti opisuju informaciju o strukturnom razredu koju pruža amplitudni spektar. Rezultati klasifikacije sekvenci u sva četiri spektralna razreda nisu dovoljno uspješni gleda li se samo amplitudni spektar. Stoga je iskoristišen skup značajki koji se koristi u aktualnoj aplikaciji za određivanje sekundarne strukture te mu je dodano prvih pet glavnih spektralnih komponenti.

Klasifikacija razreda a, b, c i d (dodano prvih pet glavnih spektralnih komponenti)					
Širina Gaussove fje	Točnost na a	Točnost na b	Točnost na c	Točnost na d	Ukupno
0.6	93%	80%	77%	73%	81%

Stvarno stanje \ Klasificirano stanje	a	b	c	d
a	410	2	5	26
b	7	352	6	78
c	14	4	266	62
d	32	43	42	324

Ukupna točnost klasifikacije je relativno 2.5% uspješnija od postojećih metoda za određivanje strukturnog razreda sekvenci.

5 Zaključak

Budući da polarnost bočnog lanca aminokiselina utječe na oblik cijelog proteinskog lanca, samim time utječe i na sekundarnu strukturu proteina. Na temelju vrijednosti indeksa hidrofobnosti aminokiselina sekvence aminokiselina transformirane su u realne diskretne signale nad kojima je potom provedena frekvencijska analiza. Spektri signala sekvenci strukturnih razreda *a* i *b* pokazali su generalni trend dominantnih spektralnih komponenti u disjunktivnim frekvencijskim područjima pa je njih najlakše bilo klasificirati, dok sva četiri razreda nije moguće uspješno klasificirati. Ipak, uspješnost aktualne metode koja se koristi za određivanje spektralnog razreda relativno je 2.5% uspješnija dodavanjem prvih pet glavnih spektralnih komponenti.